

Genetic polymorphism of HLA class I

Citation for published version (APA):

Swelsen, W. T. N. (2005). *Genetic polymorphism of HLA class I*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20050930ws>

Document status and date:

Published: 01/01/2005

DOI:

[10.26481/dis.20050930ws](https://doi.org/10.26481/dis.20050930ws)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Summary

Summary

The human leukocyte antigen (HLA) genes, encoded within the major histocompatibility complex (MHC), present the most extensive allelic polymorphism of the human genome. They are located on the short arm of chromosome 6 and divided into three regions, class I, II and III. The major classical HLA class I molecules, HLA-A, -B and -C, are expressed on all nucleated cells and platelets, whereas the class II molecules, HLA-DR, -DQ and -DP, have a more restricted expression. The essential role of the HLA molecules is the presentation of self and foreign antigens to the immune system. HLA typing is of clinical importance for solid organ transplantation, bone marrow and stem cell transplantation and disease susceptibility.

The polymorphism of HLA class I genes is mainly located in exons 2 and 3 encoding the peptide binding groove of the HLA molecules. However, more and more polymorphism is detected in the coding regions outside exons 2 and 3 and in the non-coding regions. Historically HLA polymorphism was detected by serological methods. The discovery of DNA amplification techniques introduced various molecular HLA typing methods. Sequencing of a gene is the most detailed method to type for all alleles. In addition, novel polymorphisms will be identified directly. In this thesis the polymorphism of HLA class I genes is determined for the coding regions, including exons 1 through 5 as well as for the non-coding regions, the introns.

In Chapter 1 an overview is given of the genomic organization of the MHC region and the structural features and the molecular genetics of the HLA class I molecules. Also antigen processing and presentation, extent of polymorphism, nomenclature system, clinical implications of HLA typing and different HLA typing methods, with special emphasis on sequence-based typing are discussed. The aim and scope of the thesis are explained.

Chapter 2 describes a reliable and efficient high-resolution sequence-based typing strategy for HLA-A. Exons 2 and 3 were amplified separately and sequenced heterozygously in both forward and reverse direction. Validation of the method was obtained by sequencing 11 individuals carrying alleles from all major HLA-A allele groups. Furthermore, a total of 252 unrelated individuals with at least one allele belonging to the A10 or A19 group were typed for HLA-A by this strategy. All alleles were identified correctly except A*3401. Unexpected polymorphic positions were shown in exon 3, even in individuals homozygous for A*3401. The heterozygous results in A*3401 positive individuals proved to be due to coamplification of a pseudogene, HLA-COQ or HLA-DEL, linked to A*3401. The problem was resolved by using different amplification primers for exon 3 with mismatches for the two pseudogenes.

In Chapter 3 the sequence-based typing strategy developed for HLA-B was used to type 225 unrelated individuals with at least one allele belonging to the B5/35 cross-reacting group. The heterozygous sequencing strategy included separate amplification of exons 2 and 3 using amplification primers located in introns 1, 2 and 3. The SBT strategy proved to be reliable and efficient for high resolution typing of B5/35 CREG alleles. Furthermore the exon 1, 4 and 5 sequences of 26 different B5/35 alleles were determined by allele-specific sequencing to assess the level of polymorphism. Three new alleles were detected, one by SBT of exons 2 and 3 (B*5204) and two by sequencing exons 1, 4 and 5 (B*3542 and B*510105). The study suggested that more polymorphism might be present outside exons 2 and 3 than previously thought.

Sequence-based typing methods for HLA-B are based on elucidation of exon 2 and 3 sequences. The elucidation of the exons 1, 4 and 5 sequences has led to an increase of ambiguities with alleles having identical exon 2 and 3 sequences but differences in other exons. The ambiguities can be resolved by sequencing the exon in which the difference is located. In Chapter 4, an SBT strategy for heterozygous sequencing of exons 1, 4 and 5 of HLA-B has been described. The strategy was validated against a panel of 25 individuals, carrying HLA-B alleles from 33 different allele groups and was used to resolve ambiguities in 50 individuals. It proved to be a valuable tool for resolving ambiguities of HLA-B alleles with differences in these exons, as well as for studying the polymorphism of HLA-B outside exons 2 and 3.

At present more than 600 HLA-B alleles are identified by sequence analysis. The polymorphic exon 2 and 3 sequences of all HLA-B alleles have been described. However, the exon 1, 4 and 5 sequences of a number of infrequent HLA-B alleles are unknown. In Chapter 5 the exon 1, 4 and/or 5 sequences of 39 of such B-alleles were elucidated by allele-specific sequencing. In general, these exon sequences showed identity with the majority of the known sequences from the corresponding allele groups. The polymorphism observed in these exons merely reflects the lineage specificity of HLA-B. Several alleles showed exon 1, 4 and/or 5 sequences not resembling the sequences of the corresponding allele group namely; B*4010, B*4415, B*4416 and B*5606. The deviating sequences found were either confirming the evolutionary origin of the alleles or showed new polymorphisms.

The number of HLA-B alleles is increasing continually and therefore molecular methods have to be updated regularly. In Chapter 6 an extended sequence-based typing method for subtyping of B*27 and an updated polymerase chain reaction sequence-specific primer (PCR-SSP) method to detect the presence or absence of B*27 are described. The presence of B*27 is used as a

diagnostic marker for ankylosing spondylitis. The sequences of exons 1 through 4 of 78 individuals were determined to unequivocally assign the B*27 alleles. Eleven different subtypes were detected including three new alleles: B*270504, B*2715 and B*2717.

The sequence database of HLA class I genes focuses on the coding sequences, the exons. Limited information is available on the non-coding sequences of the different HLA class I alleles. The unknown intron 1, 2 and 3 sequences of B*7301 were elucidated in Chapter 7. Overall the B*73 sequence resembles the sequence of other HLA-B alleles, although 35 differences were found with a consensus intron sequence. The insertions and deletions shown in intron 2 of B*73 were strikingly similar to the sequences of the HLA-C alleles, as was the 5' end of intron 3. Furthermore a unique deletion was observed in the middle of intron 3, not present in HLA-B or -C alleles. In Chapter 7 the reason that B*73 was not amplified with our HLA-B SBT strategy was unraveled by analysis of the intron sequences of B*7301. The heterozygous amplification primers of the HLA-B SBT protocol showed mismatches with B*73 intron sequences, causing the allelic drop-out. Correct amplification of exons 2 and 3 of B*7301 was enabled by the design of new primers in introns 2 and 3.

To assess the degree of polymorphism the intron 4 nucleotide sequences of at least one representative of each major allelic group of HLA-A, -B and -C have been determined by allele-specific sequencing in Chapter 8. The sequences revealed that the length of intron 4 varies as a result of insertions and deletions with a minimum of 93 and a maximum of 124 nucleotides. There were remarkable similarities and differences within HLA-A, -B and -C alleles, as well as between them. Both HLA-A and -B alleles can be divided into two major groups, one with a deletion and one without a deletion in intron 4. Remarkable was the deletion of 20 nucleotides in all HLA-A and -B alleles compared with HLA-C. Thirty-three different allelic groups were represented by more than one allele. The intron 4 nucleotide sequences were relatively conserved within these groups. In three allelic groups B*40, B*48 and Cw*07, differences in the intron 4 sequence were noticed between alleles. The intron 4 sequence of B*7301 resembles the sequences of other B-alleles, the deletion of 20 nucleotides is present in B*7301. Although the exon 4 and 5 sequences of B*7301 resemble the HLA-C sequences, the elucidated intron 4 sequence makes a gene conversion event of exons 4 and 5 of HLA-C for the origin of B*7301 unlikely.

Chapter 9 describes an unusual haplotype detected in a family of a Caucasian renal transplant patient. HLA analysis of ten family members in two generations demonstrated the absence of HLA-A on one of the haplotypes present in the

patient and a sibling. The absence of any serological or molecular reaction together with the results of microsatellite and fluorescent in situ hybridisation (FISH) analysis suggested a deletion of a very narrow region, encompassing the HLA-A gene. The actual presence and extent of the deletion could not be proven, however it is unlikely that the findings can be explained by another phenomenon.

Chapter 10 contains the general discussion and places the different topics studied in the previous chapters, in perspective. Polymorphism is present in exons 2 and 3, but also outside these regions both in the coding and non-coding regions. The polymorphism observed in these regions merely reflects the lineage specificity of the different HLA class I genes. The actual extent and functional role of polymorphism outside the peptide binding groove is still an issue to be investigated. However, it is unquestionable that the HLA class I genes display an extensive polymorphism and that the number of polymorphic positions will increase further in the future.

Samenvatting

Samenvatting

De HLA (human leukocyte antigen) genen zijn de meest polymorfe genen die bij de mens bekend zijn. De genen zijn gelokaliseerd op de korte arm van chromosoom 6 en liggen in een gebied dat het Major Histocompatibility Complex (MHC) genoemd wordt. De functie van HLA moleculen is de controle van zelfherkenning en dus verdediging tegen vreemde antigenen. De HLA moleculen presenteren antigenen aan het immuunsysteem. De presentatie van eigen peptiden zal, normaal gesproken, genegeerd worden door het immuunsysteem als gevolg van zelftolerantie. Echter, de presentatie van vreemde peptiden zal leiden tot de activatie van immuuncellen. HLA typering is van klinisch belang voor orgaantransplantaties, beenmerg/stamcel transplantaties en bij ziekte associaties.

Binnen het MHC kunnen drie regio's worden onderscheiden; klasse I, klasse II en klasse III. De klasse I regio bevat genen die coderen voor de zware keten van de drie klassieke HLA moleculen, HLA-A, -B en -C. De klasse II genen leveren de genprodukten voor de HLA-DR, -DQ en -DP moleculen. De klasse I en II regio's behoren tot het HLA complex. De klasse III regio bevat genen die coderen voor moleculen met verschillende functies, zoals componenten van het complement systeem en cytokines. Het onderzoek in dit proefschrift richt zich op HLA klasse I genen.

De HLA klasse I moleculen zijn glycoproteïnen die aanwezig zijn op het oppervlak van alle kernhoudende cellen en bloedplaatjes. Ze bestaan uit twee polypeptide ketens, een polymorfe MHC-gecodeerde zware keten met een molekulgewicht van 45 kD en een lichte keten van 12 kD. Beide ketens zijn non-covalent gebonden. De zware keten bestaat uit drie extracellulaire domeinen, $\alpha 1$, $\alpha 2$ en $\alpha 3$, een hydrofoob transmembraan gebied en een hydrofiële cytoplasmatische staart. De zeer polymorfe $\alpha 1$ en $\alpha 2$ domeinen vormen samen de groeve waarin peptiden worden gebonden. De lichte keten is het $\beta 2$ -microglobuline dat gecodeerd wordt door een gen op chromosoom 15.

De genen die coderen voor de zware keten van de HLA klasse I moleculen hebben een karakteristieke structuur waarbij de genetische informatie voor de verschillende domeinen van het eiwit gelokaliseerd is in verschillende exonen. De HLA-A en -C genen bevatten 8 exonen en het HLA-B gen zeven. Exon 1 codeert voor het leaderpeptide en de exonen 2, 3 en 4 voor de extracellulaire $\alpha 1$, $\alpha 2$ en $\alpha 3$ domeinen. De genetische informatie voor het transmembraan gebied is gelokaliseerd in exon 5 en voor de cytoplasmatische staart in exonen 6, 7 en/of 8.

Het extreme polymorfisme van HLA genen wordt geïllustreerd door het grote aantal allelen dat bekend is voor de verschillende loci. Het aantal allelen is de afgelopen jaren enorm gegroeid en dit zal in de toekomst waarschijnlijk blijven gebeuren. HLA-B is het meest polymorfe locus met 626 verschillende allelen. Voor HLA-A zijn 349 en voor HLA-C 182 allelen bekend. Het polymorfisme van HLA moleculen verschaft individuen de diversiteit in antigen presentatie, noodzakelijk voor populatieoverleving. Nieuwe allelen ontstaan door verandering in de base volgorde van het gen. Genetische mechanismen hierbij kunnen zijn: puntmutaties, deleties, duplicaties, recombinaties en conversies. Sommige polymorfe varianten hebben invloed op de HLA expressie. HLA allelen die niet op het oppervlak van een cel tot expressie komen worden nul-allelen genoemd.

Er bestaan verschillende technieken om HLA polymorfisme te detecteren. In het verleden werden vooral serologische methoden gebruikt. De introductie van DNA amplificatie technieken heeft geleid tot HLA typering op genetisch niveau. De drie moleculaire technieken die het meest gebruikt worden zijn, PCR-SSP (sequentie specifieke primers), PCR-SSOP (sequentie specifieke oligonucleotide probes) en PCR-SBT (sequentie gebaseerd typeren). Het typeren van HLA allelen door middel van sequentie-analyse is de meest volledige methode, waarbij bovendien nieuwe polymorfismen direct kunnen worden geïdentificeerd.

Het polymorfisme van HLA klasse I genen is voornamelijk gelokaliseerd in de exonen 2 en 3, welke coderen voor de peptide bindingsgroeve van de HLA moleculen. Steeds meer polymorfismen worden echter gedetecteerd in de coderende regio's buiten exon 2 en 3 alsook in de niet-coderende regio's. In dit proefschrift is het polymorfisme bestudeerd van HLA klasse I genen voor de coderende regio's, de exonen 1 tot en met 5 en de niet-coderende regio's, de intronen.

In hoofdstuk 1 wordt een overzicht gegeven van de genomische organisatie van het MHC gebied, de structurele kenmerken en de moleculaire genetica van de HLA klasse I moleculen. Ook worden antigeen-productie, peptide-presentatie, het uitgebreide polymorfisme, nomenclatuur, klinische implicaties van HLA typering en de verschillende HLA typeermethoden, met speciale nadruk op SBT, bediscussieerd. Vervolgens wordt het doel van dit proefschrift uiteengezet.

Hoofdstuk 2 beschrijft een betrouwbare en efficiënte hoge resolutie SBT-strategie voor HLA-A. De exonen 2 en 3 werden afzonderlijk geamplificeerd en de sequentie werd heterozygoot bepaald van zowel de coderende als de

complementaire DNA streng. De methode werd gevalideerd door sequentie-analyse van 11 individuen, positief voor allelen uit de meest voorkomende HLA-A allel groepen. Daarnaast werden 252 niet-verwante individuen, positief voor ten minste 1 allel behorend tot de A10 of A19 groep, met deze strategie getypeerd voor HLA-A. Alle allelen werden correct geïdentificeerd behalve het allel A*3401. Hierbij werden onverwachte polymorfe posities gevonden in exon 3, zelfs in individuen die homozygoot waren voor A*3401. De heterozygote resultaten in A*3401 positieve individuen waren het gevolg van co-amplificatie van een pseudogen, HLA-COQ of HLA-DEL, gekoppeld aan A*3401. Het probleem werd opgelost door het gebruik van andere amplificatie primers voor exon 3 die gemismatcht waren met de twee pseudogenen.

In hoofdstuk 3 wordt de SBT-strategie, ontwikkeld voor HLA-B, gebruikt om 225 niet-verwante individuen met ten minste 1 allel behorend tot de B5/35 kruisreagerende groep te typeren. De heterozygote sequentie-strategie bestond uit afzonderlijke amplificatie van de exonen 2 en 3, waarbij amplificatie primers gebruikt werden die gelokaliseerd waren in de intronen 1, 2 en 3. De SBT-strategie bleek betrouwbaar en efficiënt te zijn voor hoge resolutie typering van B5/35 allelen. Daarnaast werden de exon 1, 4 en 5 sequenties van 26 verschillende B5/35 allelen bestudeerd door middel van allel-specifieke sequentie-bepaling, om de mate van polymorfisme te onderzoeken. Drie nieuwe allelen werden geïdentificeerd, een met SBT van de exonen 2 en 3 (B*5204) en twee met sequentie-analyse van de exonen 1, 4 en 5 (B*3542 en B*510105). De studieresultaten suggereerden dat meer polymorfisme aanwezig is buiten de exonen 2 en 3 dan eerder werd aangenomen.

SBT-strategieën voor HLA-B zijn gebaseerd op exon 2 en 3 sequenties. De opheldering van exon 1, 4 en 5 sequenties heeft geleid tot een toename van ambiguïteiten waarbij allelen identieke exon 2 en 3 sequenties hebben, maar verschillen in andere exonen. De ambiguïteiten kunnen worden opgelost door de exon-sequentie te bepalen op de plaats waar het verschil gelokaliseerd is. In hoofdstuk 4, wordt een SBT-strategie voor de heterozygote sequentie-bepaling van exonen 1, 4 en 5 van HLA-B beschreven. De strategie werd gevalideerd met behulp van een panel van 25 individuen met HLA-B allelen uit 33 verschillende allel groepen en gebruikt om ambiguïteiten op te lossen van 50 individuen. De strategie bleek een waardevol instrument te zijn, om ambiguïteiten van HLA-B allelen met verschillen in exonen 1, 4 en 5 op te lossen alsmede voor het bestuderen van HLA-B polymorfisme buiten de exonen 2 en 3.

Op dit moment zijn met sequentie-analyse meer dan 600 HLA-B allelen geïdentificeerd. De polymorfe exon 2 en 3 sequenties van alle HLA-B allelen

zijn beschreven, maar de exon 1, 4 en 5 sequenties van een aantal niet frequent voorkomende HLA-B allelen zijn onbekend. In hoofdstuk 5 werden de exon 1, 4 en/of 5 sequenties van 39 HLA-B allelen opgehelderd door allel-specifieke sequentie-bepaling. In het algemeen kwamen deze exon-sequenties overeen met de reeds bekende sequenties van de corresponderende allel groepen. Het polymorfisme, gezien in deze exonen, weerspiegelde de afkomst van het allel. Een aantal allelen vertoonden exon 1, 4 en/of 5 sequenties die niet overeenkwamen met de sequenties van de corresponderende allel groepen namelijk: B*4010, B*4415, B*4416 en B*5606. De gevonden afwijkende sequenties bevestigden de evolutionaire afkomst van de allelen of bleken nieuwe polymorfismen te zijn.

Aangezien het aantal HLA-B allelen voortdurend toeneemt moeten moleculaire methoden regelmatig worden bijgewerkt. Hoofdstuk 6 beschrijft een uitgebreide SBT methode voor subtypering van B*27 en een bijgewerkte PCR-SSP methode om de aan- en afwezigheid van B*27 te detecteren. De aanwezigheid van B*27 wordt gebruikt als een diagnostische marker voor ankylosis spondylitis. Om ondubbelzinnig B*27 allelen te detecteren werden de exon 1 tot en met 4 sequenties bepaald van 78 individuen. Elf verschillende subtypes werden geïdentificeerd waaronder drie nieuwe allelen: B*270504, B*2715 en B*2717.

De sequentie databank van HLA klasse I genen is gericht op de coderende sequenties, de exonen. Beperkte informatie is beschikbaar voor de niet-coderende sequenties van de verschillende HLA klasse I allelen. De onbekende intron 1, 2 en 3 sequenties van B*7301 werden opgehelderd in hoofdstuk 7. Over het algemeen vertoonde de intron 1-3 sequentie van B*73 overeenkomst met de sequentie van andere HLA-B allelen, hoewel er 35 verschillen gevonden werden met de consensus intron-sequentie. De inserties en deleties van B*73 in intron 2 waren opvallend gelijk aan de sequenties van HLA-C allelen, wat ook het geval was voor de 5' uiteinde van intron 3. Daarnaast werd er een unieke deletie gevonden in het midden van intron 3, die niet aanwezig was in andere HLA-B noch in HLA-C allelen. Door de analyse van de intron-sequenties van B*7301 is de reden ontrafeld waarom B*73 niet geamplificeerd werd met onze HLA-B SBT-strategie. De heterozygote amplificatie primers bleken gemismatcht te zijn met de B*73 intron-sequenties, waardoor een allel drop-out werd veroorzaakt. Correcte amplificatie van de exonen 2 en 3 van B*7301 werd mogelijk gemaakt door nieuwe primers te ontwerpen in de intronen 2 en 3.

Om het polymorfisme in intron 4 te bestuderen, werden in hoofdstuk 8 de desbetreffende sequenties van bijna alle voorkomende allel groepen van HLA-

A, -B en -C opgehelderd. Dit werd gedaan door allel-specifieke sequentie-bepaling. De sequenties lieten zien dat de lengte van intron 4 varieert als gevolg van inserties en deleties, met een minimum van 93 en een maximum van 124 nucleotiden. Er werden opmerkelijke overeenkomsten en verschillen gevonden binnen de genen HLA-A, -B en -C alsook tussen de verschillende genen. Zowel HLA-A als -B allelen kunnen worden onderverdeeld in twee grote groepen, een met deletie en een zonder deletie in intron 4. Opvallend was de deletie van 20 nucleotiden in alle HLA-A en -B allelen vergeleken met HLA-C. Drieëndertig verschillende allel groepen waren vertegenwoordigd met meer dan één allel. De intron 4 nucleotide sequenties waren betrekkelijk geconserveerd binnen deze groepen, behalve voor B*40, B*48 en Cw*07. De intron 4 sequentie van B*7301 vertoonde overeenkomst met de sequenties van andere B-allelen want de deletie van 20 nucleotiden was hier aanwezig. Hoewel de exon 4 en 5 sequenties van B*7301 overeenkomsten vertonen met C-allelen, maakte de opgehelderde intron 4 sequentie een gen conversie van exonen 4 en 5 van HLA-C als herkomst voor B*7301 onwaarschijnlijk.

Hoofdstuk 9 beschrijft een ongewoon haplotype gedetecteerd in de familie van een Kaukasische niertransplantatie patiënt. HLA analyse van 10 familieleden in twee generaties demonstreerde de afwezigheid van HLA-A op een van de haplotypen aanwezig in de patiënt en zijn zuster. De afwezigheid van enige serologische of moleculaire reactie samen met de resultaten van microsatelliet en fluorescentie in situ hybridisatie (FISH) analyses suggereerden een deletie van een klein gebied, waarbinnen het HLA-A gen gelegen is. De feitelijke aanwezigheid en grootte van de deletie konden niet definitief worden bewezen, het is echter onwaarschijnlijk dat de bevindingen verklaard kunnen worden door een ander fenomeen.

Hoofdstuk 10 bevat de algemene discussie en plaatst de verschillende onderwerpen, bestudeerd in eerdere hoofdstukken, in perspectief. Polymorfisme is aanwezig in exonen 2 en 3, maar ook buiten deze regio's zowel in coderende als niet-coderende gebieden. Het polymorfisme dat gezien wordt in deze regio's is vaak een weerspiegeling van de groep-specifieke afkomst van de verschillende HLA klasse I allelen. De werkelijke aard en functionele rol van polymorfisme buiten de peptide bindingsgroeve zullen nog verder onderzocht moeten worden. Echter, het is duidelijk dat de HLA klasse I genen een uitgebreid polymorfisme vertonen en dat aan de groei van het aantal polymorfe posities voorlopig geen eind lijkt te komen.